UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/662,550	09/15/2003	Eric Cosatto	2000-0042Con	2283

83224 7590 08/07/2009

AT & T LEGAL DEPARTMENT - NDQ
ATTN: PATENT DOCKETING
ONE AT & T WAY, ROOM 2A-207
BEDMINSTER, NJ 07921

EXAMINER	
HAJNIK, DANIEL F	

ART UNIT	PAPER NUMBER
2628	

MAIL DATE	DELIVERY MODE
08/07/2009	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/662,550	COSATTO ET AL.	
	Examiner	Art Unit	
	DANIEL F. HAJNIK	2628	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *28 May 2009*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *22-25,27-32,34 and 35* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *22-25,27 29-32 and 34* is/are rejected.

7)☒ Claim(s) *28 and 35* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *15 September 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

Response to Amendment

The affidavit filed on 5/28/2009 under 37 CFR 1.131 is sufficient to overcome the 35 USC 102(a) prior art rejection based upon the Cox reference (NPL Doc, "Speech and language processing for next-millennium communications services").

Claim Rejections - 35 USC § 103

1. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

2. Claims 22-25, 27, 29-32, and 34 are rejected under 35 U.S.C. 103(a) as being unpatentable over Ezzat et al. (NPL document, "Visual Speech Synthesis by Morphing Visemes" the 1999 version, herein referred to as "Ezzat") in view of Jiang et al. (NPL document, "Visual Speech Analysis with Application to Mandarin Speech Training", herein referred to as "Jiang") in view of Hunt (NPL Doc, "Unit selection in a concatenative speech synthesis system using a large speech database").

As per claim 22, Ezzat teaches the claimed:

selecting via a processor candidate image samples utilizing the target feature vector to generate a photo-realistic animation of the object, wherein generating the photo- realistic animation of the object *(in the abstract, "we are able to synchronize the visual speech stream with the audio speech stream, and hence give the impression of a photorealistic talking face" and the middle of the 1st col on page 6 "there are many intermediate frames that lie between the chosen viseme images ... Consequently, we compute a series of consecutive optical flowvectors between each intermediate image and its successor, and concatenate them all into one large flow vector that defines the global transformation between the chosen visemes"; in this case, the visemes represent a candidate image samples that can be use to describe a particular sound and the flowvectors which contain visual and sound features are used in conjunction with the visemes);.*

Ezzat does not explicitly teach the claimed the remaining claim limitations.

Jiang teaches the claimed:

obtaining, for each frame in a plurality of N frames of an object animation, a target feature vector comprising visual features and non-visual features associated with the object animation *(in the abstract, "At each frame, region of interest is identified and key information is extracted. The preprocessed acoustic and visual information are then fed into a modular TDNN and combined for visual speech analysis" where the acoustic information represents a non-visual feature; also see pg. 114, 4.2 Acoustic and Visual Input Representation, 1st paragraph, "For acoustic data representation, we have followed the well-established approach to apply FFT on the Hamming windowed speech data to get 16 Melscale Fourier coefficients as input to the Acoustic input Layer. For visual data representation, we have performed the lip-tracking and*

feature points extraction task by applying our 2D multi-state lip shape model ... The extracted feature vectors are then fed to the Visual Input Layer")

It would have been obvious to one of ordinary skill in the art at the time of invention to combine Ezzat with Jiang in order to provide more analysis data for the speech synthesis process. Ezzat is modified by Jiang by applying the target feature vectors to the phonemes and visemes as used in Ezzat for each frame. For example, the target feature vector may be used in the output frames (labeled "Video") shown in figure 1 of Ezzat.

The combination of Ezzat and Hunt together suggest the claimed:

Selecting via a processor candidate image samples ... using an audio/video unit selection process in which a longest possible candidate image sample is selected *(Hunt: on page 1 under section 2, "Unit Selection" and Hunt in the 1st paragraph under section 2.2, "Thus, each target phoneme and each candidate in the synthesis database is characterized by a multidimensional feature vector"; Hunt teaches of picking the longest possible candidate sample by using a concatenation costs, i.e. see the bottom of the 2nd col on page 1 of Hunt: "The concatenation cost ... is an estimate of the quality of a join between consecutive units" where a candidate sample that is longer will have a lower concatenation costs because the transition between phonemes is smoother or more natural; thus the algorithm favors that longest available candidate samples).*

Hunt alone does not teach all the features related to the unit selection process because Hunt's unit selection is for phonemes; however when combined with Ezzat the claimed feature above is suggested by the combination of references.

In particular, Ezzat teaches of associating viseme images with phonemes on page 3 in the 2nd and 3rd paragraphs under section 4, i.e. Ezzat states: *"A one-to-one mapping between phonemes and visemes thus ensures that a unique viseme image is associated with each phoneme label."* When one of ordinary skill in the art applies the unit selection process in Hunt with the phoneme-viseme relationship in Ezzat, the combination acts as a unit selection process for the viseme images associated with the phonemes. One of ordinary skill in the art would modify Ezzat to achieve the claimed feature by substituting the phonemes as used in Ezzat with the phonemes as used in Hunt. Thus, when Hunt performs their unit selection process the visemes image will also perform in the unit selection process because there is a one-to-one relationship between them.

It would have been obvious to one of ordinary skill in the art at the time of invention to combine Ezzat, Jiang, and Hunt. One advantage is that using the unit selection process of Hunt results in "to produce a natural realisation of a target phoneme sequence" (see the abstract of Hunt). Further, one of ordinary skill in the art using the system of Ezzat may look to incorporate the concatenation costs of Hunt because the Ezzat refers to concatenation in their Appendix A on page 11. Thus, one of ordinary skill in the art may examiner other concatenation references such as Hunt to enhance Ezzat.

As per claim 23, this claim is similar in scope to limitations recited in claim 22, and thus is rejected under the same rationale.

As per claim 24, Ezzat teaches the claimed:

24. The method of claim 22, wherein selecting candidate image samples further comprises for each frame in the plurality of N frames of the animation *(middle of the 1st col on page 3, "From an animator's perspective, the visemes in our model represent keyframes")*, selecting candidate image samples associated with the object animation using a comparison of a combination of visual features and non-visual features with the target feature vector *(page 3, 2nd paragraph in the 1st col: "we determine the appropriate sequence of viseme morphs to make, as well as the rate of the transformations by utilizing the output of the natural language processing unit"; Bottom of the 2nd col on page 8 and top of the next page, "the lip-sync module creates the video stream, which is composed of a sequence of frames which sample the chosen viseme transitions" where the viseme transitions are visual features; 3rd paragraph under section 7, " The lip-sync module loads the appropriate viseme transitions into the viseme transition stream by examining the audio diphones" where this audio is a non-visual feature).*

In order to determine the appropriate sequence, the system performs a comparison of visual and non-visual features with a given target vector in order to produce the output as stated. In this case, the visual features are the video images of facial movements of the mouth (for example see figures 4 and 5). The non-visual features are the sounds that correspond to the facial movements (for example what sound a given syllable makes such as a phoneme, see figure 3). Further, this construction process of an appropriate sequence of viseme morphs would require selecting candidate image samples where these samples could be used to transition between through transformation.

As per claim 25, Ezzat teaches the claimed:

24. The method of claim 24, further comprising compiling the selected image sample candidates to form a photo-realistic animation *(in the abstract, "A complete visual utterance is constructed by concatenating viseme transitions. Finally, phoneme and timing information extracted from a text-to-speech synthesizer is exploited to determine which viseme transitions to use ... and hence give the impression of a photorealistic talking face").*

As per claim 27, Ezzat teaches the claimed:

27. The method of claim 22, further comprising:

creating a first database of image samples showing an object in a plurality of appearances *(by teaching of recording and collecting one image per English phoneme- 2nd paragraph under section 3 under "Corpus and Viseme Acquisition", also see figure 2)*;

creating a second database of the visual features for each image sample of the object *(in figure 8 where video frames for each viseme or image sample are created; also visual features of the optical flow correspondence is stored in the second data, 3rd paragraph under section 7)*; and

creating a third database of the non-visual features of the object in each image sample *(in figure 8 where diphone data or audio data that is stored in a database for each viseme or image sample).*

As per claim 29, Ezzat teaches the claimed:

29. The method of claim 27, wherein the animation is a talking- head animation *(in the abstract, "we are able to synchronize the visual speech stream with the audio speech stream, and hence*

give the impression of a photorealistic talking face"), the first database stores sample images of a face that speaks *(by teaching of recording and collecting one image per English phoneme- 2nd paragraph under section 3 titled: "Corpus and Viseme Acquisition", also see figure 2)*, the second database stores associated facial visual features *(in figure 8 where video frames for each viseme or image sample are created; also visual features of the optical flow correspondence is stored in the second data, 3rd paragraph under section 7)* and the third database stores acoustic information for each frame in the form of phonemes *(in figure 8 where diphone data or audio data that is stored in a database for each viseme or image sample).*

As per claims 30-32, these claims are similar in scope to limitations recited in claims 22, 24 and 25, respectively, and thus are rejected under the same rationale.

As per claim 34, this claim is similar in scope to limitations recited in claim 27, and thus is rejected under the same rationale.

Allowable Subject Matter

Claims 28 and 35 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

The following is a statement of reasons for the indication of potentially allowable subject matter: the cited prior art does not disclose or render obvious the combination of elements recited in the claims as whole. Specifically, the cited prior art fails to disclose or render obvious the

following limitations: "selecting … candidate image samples … using an audio/video unit selection process in which a longest possible candidate image is selected" in combination with the first, second, and third database in combination with the "calculation ... a concatenation cost from a combination of visual features" in combination with "performing a Viterbi search ... through the candidates".

Response to Arguments

The prior 35 USC 101 rejection of claims 30-32, 34, and 35 has been withdrawn in view of the respective claim amendments.

Applicant's arguments have also been considered but are moot in view of the new ground(s) of rejection. In particular, the Cox reference (dated Aug 2000) has been removed from the prior art rejections of the claims. Further, the reference of Hunt is now relied upon for teaching some of the claimed features (i.e. unit selection).

Also, please note that in this office action, the Ezzat NPL reference (dated 1999) relied upon is an earlier version of another Ezzat NPL reference (dated 2000) previously relied upon. A copy of the Ezzat 1999 version is attached with this office action.

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to DANIEL F. HAJNIK whose telephone number is (571)272-7642. The examiner can normally be reached on Mon-Fri (8:30A-5:00P).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Ulka Chauhan can be reached on (571) 272-7782. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Daniel F Hajnik/
Examiner, Art Unit 2628

/Peter-Anthony Pappas/
Primary Examiner, Art Unit 2628